Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Sound and Vibration 275 (2004) 331–349
JOURNAL OF SOUND AND VIBRATION
www.elsevier.com/locate/jsvi

On the optimality during parameter identification

M. Boltežar*, N. Jakšić

University of Ljubljana, Faculty of Mechanical Engineering, Aškerčeva 6, 1000, Ljubljana, Slovenia

Received 10 February 2003; accepted 17 June 2003

Abstract

In this article the relations between the mathematical theory of function approximation, the maximum-likelihood method, the measures of optimality and the identification of parameters are presented. The estimator of the maximum likelihood for uniform noise is introduced on the basis of the generalized Cauchy probability density function (p.d.f.). The measures of optimality based on the maximum-likelihood method for the random variables with Gauss, Cauchy, Laplace and uniform p.d.f.s are presented. The theoretical statements are illustrated with a numerical experiment concerning the optimal parameter identification on the free-damped-noisy response of a single-degree-of-freedom system. The different types of noise and the different levels of the responses' noisiness were used.

1. Introduction

The optimization process, in our case, the parameter identification [1,2], is always governed by the search for an objective function, which is a mathematical means of carrying out the optimization. The optimization is made up of the data (measurements), the model (which is fitted to the data) and the measure that tells us how the model should be fitted to the data.

The mathematical theory of function approximation states that the condition of the optimum approximation is the minimum distance between the function and its approximation [3]. The measure for distance is not defined by the theory of approximation. However, the maximum-likelihood method helps us with the optimum expression for the distance [4].

The distance between measurements, which have a random spread around the true values, and the model is minimized by minimizing the objective function. Hence, in this case the random variable with a certain probability density function (p.d.f.) has to be minimized. The optimal expression for the distance can be deduced from the p.d.f. of the random variable with the help of

*Corresponding author. Tel.: +386-1-4771-608; fax: +386-1-2518-567.
E-mail address: miha.boltezar@fs.uni-lj.si (M. Boltežar).

0022-460X/$ - see front matter
doi:10.1016/j.jsv.2003.06.026

the maximum-likelihood method. As a result, the objective function can be formed. The restriction placed upon the p.d.f. by the maximum-likelihood method is that the inverse of function's values must exist.

This study introduces a generalized Cauchy p.d.f., which enables deduction of a maximum-likelihood estimator for uniform noise. It also introduces a more generalized instantaneous measure of the noise content in the signal, $FSNR(t)$, which can be used when dealing with Gaussian and uniform noise, as well as when dealing with noise that has a symmetrical-α-stable probability distribution.

The paper brings together some different mathematical and statistical methods in a way that is clear and applies them to the parameter identification. It also applies some probability distributions, e.g., symmetrical-α-stable and Cauchy, that have been used recently in statistics, signal processing and electro-engineering fields, but much less so in papers dealing with mechanical engineering issues.

First, a short recapitulation of the function approximation theory is presented in the theoretical part. Then the maximum-likelihood method is summarized, and the maximum-likelihood estimators for different probability density functions are defined. The different noise-level measures in a signal are also presented.

Theoretical statements are clearly presented during the numerical experiment, by parameter identification on the single-degree-of-freedom responses with different levels of noise contamination and different noise types.

2. Theory

2.1. Function approximation

Let f be defined in the interval $[a,b]\subset\mathbb{R}$. A real non-negative number $\|f\|$, called the norm, can be assigned to each function f. $\|f\|$ should satisfy the axioms of the norm:

$$\begin{aligned}
&(1)\quad \|f\| \geqslant 0,\\
&(2)\quad \|f\| = 0 \quad \text{if and only if} \quad f = 0 \text{ everywhere in } [a,b],\\
&(3)\quad \|\alpha f\| = |\alpha|\,\|f\|, \quad \text{for all } \alpha\in\mathbb{R},\\
&(4)\quad \|f+g\| \leqslant \|f\| + \|g\|.
\end{aligned} \tag{1}$$

The distance between two functions f_1 and f_2, defined in $[a,b]$, is defined as

$$\varrho(f_1,f_2) = \|f_1 - f_2\|, \tag{2}$$

which satisfies the distance axioms:

$$\begin{aligned}
&(1)\quad \varrho(f_1,f_2)\geqslant 0,\\
&(2)\quad \varrho(f_1,f_2) = 0 \quad \text{if and only if} \quad f_1(t) = f_2(t) \text{ everywhere in } [a,b],\\
&(3)\quad \varrho(f_1,f_2) = \varrho(f_2,f_1),\\
&(4)\quad \varrho(f_1,f_3)\leqslant \varrho(f_1,f_2) + \varrho(f_2,f_3).
\end{aligned} \tag{3}$$

Let f be a discrete function defined in the interval $[a,b] \subset \mathbb{R}$ on a finite number of points (t_i, f_i), where $f_i = f(t_i)$. The function f is approximated with the function $\Phi \in \mathbb{R}$, which is dependent on the argument t and on M parameters, $a_1, a_2, \ldots, a_M$.

$$\Phi(t) = \Phi(t, a_1, a_2, \ldots, a_M) = \Phi(t; \mathbf{a}). \tag{4}$$

The approximation problem can be defined as follows [3]: for the given function f in $[a,b]$, find the function Φ of the shape (4) that gives a minimum distance $\|f - \Phi\|$. Such an approximation function is called the best approximation of the function f for a given set of parameters **a** and for the chosen norm. Hence, the parameters **a** must be determined in a way that the distance between the functions f and Φ is minimal.

2.2. The maximum-likelihood method

The maximum-likelihood method helps us to determine the most appropriate norm when dealing with the approximation of a function [4]. The approximation function Φ is often called a model.

Suppose—as the opposite of the approximation problem—that the N data points (x_i, y_i), $i = 1, \ldots, N$, are fitted to a model that has M, $M < N$, adjustable parameters $a_1, \ldots, a_M$. For a particular set of parameters a question is raised: what is the probability that this data set could have occurred? If the model Φ takes on continuous values, the probability will always be zero, unless some fixed Δf on each data point (f_i, t_i) is considered. If the probability of obtaining the data set is infinitesimally small, then the parameters **a** under consideration are unlikely to be correct. Conversely, the data set should not be too improbable for a correct choice of parameters. The values of the parameters **a** are found by a maximization of the probability of the data set. This form of parameter estimation is the maximum-likelihood estimation.

Suppose that each data point (f_i, t_i) has a measurement error that is independently random and distributed with a given distribution, Eq. (5), around the model $f(t_i)$.

$$g(x_i) = \mathrm{e}^{-r(x_i)}, \tag{5}$$

where x_i denotes a random variable (measurement error) at data point (f_i, t_i). The probability P of a given set of data is the product of the probability of each data point:

$$P \propto \prod_{i=1}^{N} \{\mathrm{e}^{[-r(x_i)]} \Delta f\}, \tag{6}$$

where

$$x_i = f_i - f(t_i; \mathbf{a}). \tag{7}$$

Maximizing Eq. (6) is equivalent to maximizing its logarithm, or minimizing the negative of its logarithm, Eq. (8).

$$\Psi^{\star} = \left[\sum_{i=1}^{N} r(x_i)\right] - N \ln \Delta f. \tag{8}$$

Since N and Δf are constants, minimizing Eq. (8) is equivalent to minimizing Eq. (9).

$$\Psi = \sum_{i=1}^{N} r(x_i). \tag{9}$$

The function r in Eq. (9) stands for the exponent of the presumed p.d.f., Eq. (5). It can be seen from Eqs. (5) and (9) that the inverse values of the p.d.f. must exist in the interval $(-\infty, +\infty)$. The function Ψ, Eq. (9), is usually called an objective function or a merit function, or even a cost function, of the optimization problem.

2.3. Probability distributions and the maximum-likelihood estimator

The various probability distributions were taken into account, these include: Gaussian (normal), Laplace (double-sided exponential), Cauchy (Lorentzian), Symmetric-α-Stable and uniform. The estimate of a random variable is denoted as a hat (^).

2.3.1. The Gaussian probability distribution

The characteristic function of the Gaussian probability distribution [5] can be written as

$$\phi(\tau) = \mathrm{e}^{\mathrm{i}\mu\tau - (\sigma^2\tau^2)/2}, \tag{10}$$

where $\mathrm{i} = \sqrt{-1}$. The mean is denoted by μ, $\hat{\mu} = (1/N)\sum_{i=1}^{N} x_i$, and variance by σ^2, $\hat{\sigma}^2 = (1/(N-1))\sum_{i=1}^{N}(x_i - \hat{\mu})^2$. The p.d.f. is defined as

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-(x-\mu)^2/(2\sigma^2)} \tag{11}$$

and its exponent as

$$r(x) = \frac{(x-\mu)^2}{2\sigma^2}. \tag{12}$$

Hence, the objective function Ψ is formed as

$$\Psi = \sum_{i=1}^{N} r(x_i) = \sum_{i=1}^{N}\left[\frac{f_i - f(t_i;\mathbf{a}) - \mu}{2\sigma^2}\right]^2. \tag{13}$$

The denominator in Eq. (13) is constant. Hence, it does not contribute to the minimization and can be neglected. Supposing that $\mu = 0$ then a new objective function is formed, Eq. (14).

$$\Psi = \sum_{i=1}^{N} [f_i - f(t_i;\mathbf{a})]^2. \tag{14}$$

This is the well-known least-mean-squared-error (LMSE) method, where the L_2 approximation norm is used.

2.3.2. The Laplace probability distribution

The characteristic function of the Laplace probability distribution [5] can be written as

$$\phi(\tau) = \frac{e^{i\mu\tau}}{1 + \eta^2\tau^2} \tag{15}$$

the parameter μ stands for the mean, $\hat{\mu} = (1/N)\sum_{i=1}^{N} x_i$, and η for the spread around the mean value, $2\eta^2 = \sigma^2$, $\hat{\sigma}^2 = (1/(N-1))\sum_{i=1}^{N}(x_i - \hat{\mu})^2$. The p.d.f. is defined as

$$p(x) = \frac{1}{2\eta} e^{-|x-\mu|/\eta} \tag{16}$$

and its exponent as

$$r(x) = \frac{|x-\mu|}{\eta}. \tag{17}$$

Thus the objective function Ψ is formed as

$$\Psi = \sum_{i=1}^{N} r(x_i) = \sum_{i=1}^{N} \left| \frac{f_i - f(t_i; \mathbf{a}) - \mu}{\eta} \right|. \tag{18}$$

It can be seen again that the denominator in Eq. (18) is constant and because of that it can be neglected. Supposing that $\mu = 0$ then a new objective function is formed, Eq. (19).

$$\Psi = \sum_{i=1}^{N} |f_i - f(t_i; \mathbf{a})|. \tag{19}$$

In this way the least-absolute-deviation (LAD) method is formed, where the L_1 approximation norm is used.

2.3.3. The Cauchy probability distribution

The characteristic function of the Cauchy probability distribution [5] can be written as

$$\phi(\tau) = e^{im\tau - \gamma|\tau|}. \tag{20}$$

The parameter m stands for the location parameter (median), and the parameter γ for the dispersion, $\gamma > 0$. The p.d.f. is defined as

$$p(x) = \frac{2}{\pi b} \frac{1}{1 + (2(x-m)/b)^2} \tag{21}$$

where b stands for the width of the p.d.f. at its half-height and at the same time $b = 2\gamma$. The dispersion γ can be calculated using an algorithm (29), taking into account that from the 2nd step one should jump directly to the 6th step of the algorithm and consider $\alpha = 1$.

The p.d.f. can be defined as

$$p(x) = e^{\ln[(2/(\pi b))(1/(1+(2(x-m)/b)^2))]} = e^{-\ln[(\pi b/2)(1+(2(x-m)/b)^2)]} \tag{22}$$

and its exponent as

$$r(x) = \ln\left[\frac{\pi b}{2}\left(1 + \left(\frac{2}{b}(x-m)\right)^2\right)\right] \tag{23}$$

and finally, the objective function Ψ is formed as

$$\Psi = \sum_{i=1}^{N} r(x_i) = \sum_{i=1}^{N} \ln\left[\frac{\pi b}{2}\left(1 + \left(\frac{2}{b}(f_i - f(t_i; \mathbf{a}) - m)\right)^2\right)\right]. \tag{24}$$

The fraction $\pi b/2$ does not affect the minimization and can be neglected. If the zero-location parameter value is considered ($m = 0$) then a new objective function can be rewritten as

$$\Psi = \sum_{i=1}^{N} \ln\left[1 + \left(\frac{2}{b}(f_i - f(t_i; \mathbf{a}))\right)^2\right]. \tag{25}$$

The norm of the approximation is given by the sum of the logarithms. The minimization of the function (25) is be denoted with the MLC (the maximum-likelihood estimator for the Cauchy probability distribution).

2.3.4. The symmetrical-α-stable probability distribution

The characteristic function of the symmetric-α-stable (SαS) probability distribution [6,7] can be written as

$$\phi(\tau) = \mathrm{e}^{\mathrm{i}m\tau - \gamma|\tau|^{\alpha}}, \tag{26}$$

where m denotes the location parameter (median), γ stands for the dispersion, $\gamma > 0$, and α is the characteristic exponent, $0 < \alpha \leqslant 2$. In the case of $\alpha = 2$, the distribution is Gaussian. If $\alpha = 1$, then it is the Cauchy probability distribution. There exists no closed form for p.d.f.s, apart from the values of $\alpha = 1$ and $\alpha = 2$ [6].

If X is an α-stable random variable and if $0 < \alpha < 2$, then

$$\begin{aligned} &\mathrm{E}\{|X|^p\} = \infty, \quad p \geqslant \alpha, \\ &\mathrm{E}\{|X|^p\} < \infty, \quad 0 \leqslant p < \alpha \end{aligned} \tag{27}$$

and if $\alpha = 2$, then

$$\mathrm{E}\{|X|^p\} < \infty, \quad p \geqslant 0, \tag{28}$$

where $p \in \mathbb{R}$, $\mathrm{E}\{\cdot\}$ is the mathematical expectation. Hence for $0 < \alpha \leqslant 1$, symmetric-α-stable distributions have no finite first- or higher-order moments; for $1 < \alpha < 2$ they have finite first order moments and all of the fractional moments of order p, where $p < \alpha$; for $\alpha = 2$, all moments exist. In particular, all non-Gaussian SαS distributions have infinite variance [6].

The parameters γ (dispersion) and α (characteristic exponent) of a symmetric-α-stable random variable X can be computed with the following algorithm [7]:

Step 1

$$y_i = \ln(|X_i|),$$

Step 2

$$\mu_y = (1/N) \sum_{i=1}^{N} y_i,$$

Step 3

$$\sigma_y^2 = (1/(N-1)) \sum_{i=1}^{N} (y_i - \mu_y)^2,$$

Step 4

$$\alpha = \sqrt{(6/\pi^2)\sigma_y^2 - 0.5},$$

Step 5

$$C_e = 0.57721\ldots; \quad \text{Euler constant},$$

Step 6

$$\gamma = \mathrm{e}^{\mu_y - C_e(1/\alpha - 1)\alpha}. \tag{29}$$

For Gaussian processes, the most commonly used criterion for the best estimate is the LMSE criterion. With this criterion the best estimate is the one that minimizes the variance of the estimation error. If the process is Gaussian, it can be shown that this criterion also minimizes the probability of large estimation errors. For SαS processes with $\alpha < 2$, the LMSE criterion is no longer appropriate, due to the lack of finite variance. But the concept of the LMSE criterion can be easily generalized to stable processes. The minimum dispersion (MD) criterion is used. Under the MD criterion, the best estimate of a SαS random variable is the one that minimizes the dispersion of the estimation error. The dispersion of the α stable random variable plays an analogous role of variance. It can be shown that minimizing the dispersion is also equivalent to minimizing the probability of large estimation errors. For more detailed information the reader should refer to [6].

One can obtain an object function applying the minimal dispersion method, Eq. (30).

$$\Psi = \sum_{i=1}^{N} |f_i - f(t_i; \mathbf{a})|^p, \quad p < \alpha. \tag{30}$$

This is the least-mean P-norm (LPN) method. If $p = 1$, this is the LAD method, and if $p = 2$, this is the LMSE method.

2.3.5. *The uniform probability distribution*

The characteristic function of the uniform probability distribution in the interval $[a_0, a_1]$ can be written [5] as

$$\phi(t) = \frac{2}{(a_1 - a_0)\tau} \sin\left[\tfrac{1}{2}(a_1 - a_0)\,\tau \mathrm{e}^{\mathrm{i}(a_0 + a_1)\tau/2}\right]. \tag{31}$$

The p.d.f. is defined as

$$p_u(x) = \begin{cases} \dfrac{1}{a_1 - a_0}, & x \in (a_0, a_1), \\ 0, & x \notin (a_0, a_1). \end{cases} \tag{32}$$

The p.d.f. with zero mean and with a width at half-height equal to b is redefined as

$$p_c(x) = \begin{cases} \dfrac{1}{b}, & x \in \left(-\dfrac{b}{2}, \dfrac{b}{2}\right), \\ 0, & x \notin \left(-\dfrac{b}{2}, \dfrac{b}{2}\right). \end{cases} \tag{33}$$

This can be shown to have a zero first moment, $\mu = 0$, and a variance of $\sigma^2 = b^2/12$.

There is a problem with the $p_c(x)$ p.d.f. when constructing the maximum-likelihood estimator. The inverse values of the function $p_c(x)$ does not exist outside the interval $(-b/2, b/2)$.

As a result, it is necessary to find a function that would approximately represent the uniform p.d.f., Eq. (33), and for which the inverse function's values exist in the interval $(-\infty, +\infty)$. The generalized Cauchy p.d.f. is introduced for this purpose; it leads to the function (33), if subjected to the limit process $\lim_{n\to\infty}$.

The generalized Cauchy p.d.f. with zero mean (location parameter) of the random variable x is defined as

$$p(x) = \frac{2n\sin(\pi/(2n))}{\pi b}\frac{1}{1+(2x/b)^{2n}}, \quad n \geqslant 1, \tag{34}$$

where n denotes an integer exponent greater than 0, and b is the width of the function at its half-height. Some important properties of (34) are as follows:

$$\begin{aligned} &(1) \quad P = \int_{-\infty}^{+\infty} p(x)\,\mathrm{d}x = 1, \\ &(2) \quad m_1 = \mu = \int_{-\infty}^{+\infty} xp(x)\,\mathrm{d}x = \begin{cases} \infty, & n = 1, \\ 0, & n > 1, \end{cases} \\ &(3) \quad m_2 = \sigma^2 = \int_{-\infty}^{+\infty} x^2 p(x)\,\mathrm{d}x = \begin{cases} \infty, & n = 1, \\ \dfrac{b^2}{4 + 8\cos(\pi/n)}, & n > 1, \end{cases} \\ &(4) \quad \lim_{n\to\infty} m_2 = b^2/12 = \sigma^2_{p_c}, \\ &(5) \quad \lim_{n\to\infty} p(0) = 1/b = p_c(0), \end{aligned} \tag{35}$$

where P denotes probability, $m_1 = \mu$ denotes the first moment (mean), m_2 denotes the second moment and σ^2 denotes the second central moment (variance). Graphs of the function $p(x)$, Eq. (34), at $b = 1$ and $n = 1$, 2, 3, 4, 10 and 50 are shown in Fig. 1.

The generalized Cauchy p.d.f., as well as the Cauchy p.d.f., can also be written as

$$p(x) = \mathrm{e}^{\ln\left[(2n\sin(\pi/(2n))/(\pi b))\,(1/(1+(2(x-m)/b)^{2n}))\right]} = \mathrm{e}^{-\ln\left[(\pi b/(2n\sin(\pi/(2n))))(1+(2(x-m)/b)^{2n})\right]}, \quad n \geqslant 1, \tag{36}$$

where m denotes the mean, $\hat{m} = (1/N)\sum_{i=1}^{N} x_i$, if $n > 1$, and the location parameter if $n = 1$. The parameter b stands for the spread around m, $b^2 = \sigma^2(4 + 8\cos(\pi/n))$ and $\hat{\sigma}^2 = (1/(N-1))\sum_{i=1}^{N}(x_i - \hat{m})^2$, if $n > 1$ and $b = 2\gamma$ if $n = 1$. The exponent of the p.d.f. can be written as

$$r(x) = \ln\left[\frac{\pi b}{2n\sin(\pi/(2n))}(1 + (2(x-m)/b)^{2n})\right], \quad n \geqslant 1 \tag{37}$$

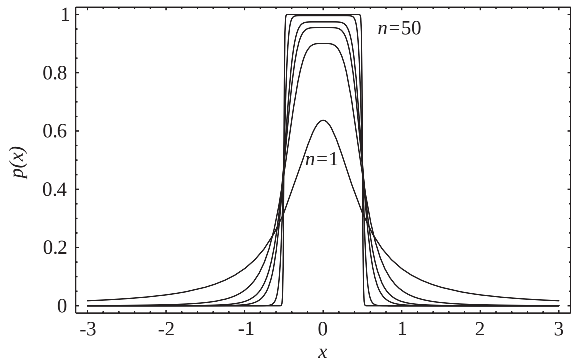

Fig. 1. Plots of the generalized Cauchy function for $b = 1$ and $n = 1, 2, 3, 4, 10$ and 50, Eq. (34).

and thus the objective function Ψ is formed as

$$\Psi = \sum_{i=1}^{N} r(z_i) = \sum_{i=1}^{N} \ln\left[\frac{\pi b}{2n\sin(\pi/(2n))}(1 + (2(f_i - f(t_i;\mathbf{a}) - m)/b)^{2n})\right], \quad n \geqslant 1. \tag{38}$$

The fraction $\pi b/[2n\sin(\pi/(2n))]$ does not affect the minimization and can be neglected. If the zero mean (location parameter) value is considered ($m = 0$) then a new objective function can be rewritten as

$$\Psi = \sum_{i=1}^{N} \ln\left[1 + (2(f_i - f(t_i;\mathbf{a}))/b)^{2n}\right], \quad n \geqslant 1. \tag{39}$$

The norm of the approximation is given by the sum of the logarithms. The minimization of the function (39) is denoted by MLU (the maximum-likelihood estimator for the uniform probability distribution). If $n = 1$ then the MLU becomes the MLC.

2.4. Measures for the noise level in a signal

A popular measure for the noise level in a signal is the signal-to-noise ratio (SNR) [1], which is defined as the power of the signal to the power of the noise in dB.

$$SNR = 10\log\frac{\mathrm{E}\{s^2\}}{\mathrm{E}\{e^2\}} \approx 10\log\frac{\mathrm{var}(s)}{\mathrm{var}(e)}, \tag{40}$$

where $\mathrm{E}\{\cdot\}$ stands for the mathematical expectation, s for signal (noise free) and e for noise. There is a different measure for asymptotical signals, like the response of the single-degree-of-freedom linear system, Fig. 2, called the instantaneous signal-to-noise ratio $SNR(t)$ [8], Eq. (41):

$$SNR(t) = 10\log\frac{X_0^2 \mathrm{e}^{-2\delta\omega_o t}}{2\,\mathrm{var}(e)}, \tag{41}$$

where X_0 stands for the initial amplitude, δ for the viscous damping ratio and ω_o for the natural frequency of the undamped system.

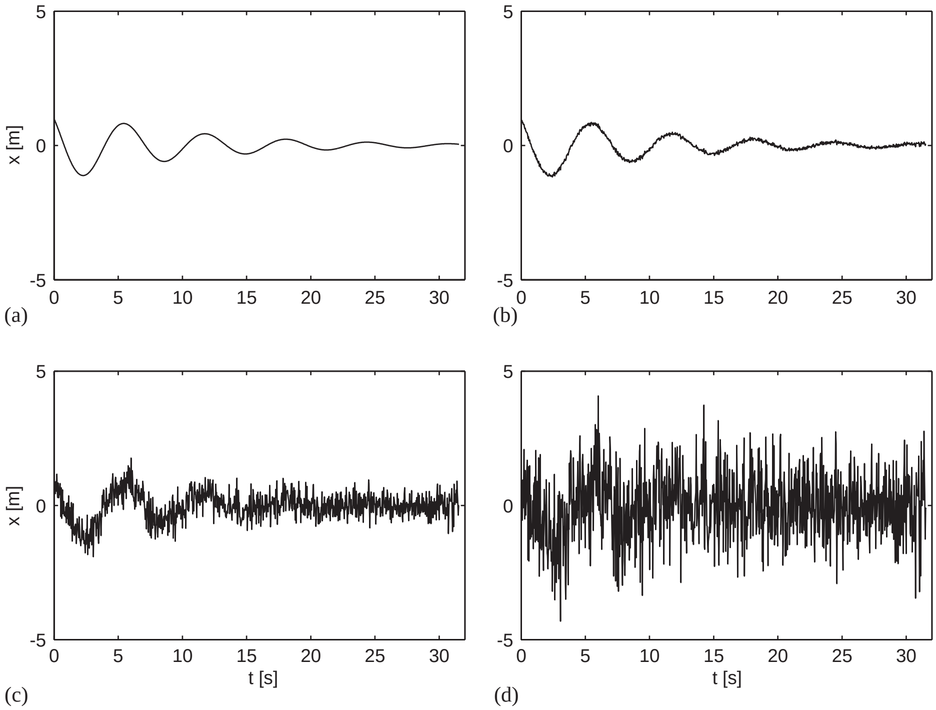

Fig. 2. System's response: (a) without added noise, (b) added Gaussian noise $FSNR = 10$ dB, (c) added Gaussian noise $FSNR = 0$ dB and (d) added Gaussian noise $FSNR = -5$ dB.

Table 1
The comparison between *SNR* and *FSNR*

FSNR	$p = 0.5$	0.9	1.0	1.5	2.0
	$\approx SNR$				
−5	−20	−11	−10	−7	−5
0	0	0	0	0	0
5	20	11	10	7	5
10	40	22	20	13	10
20	80	44	40	27	20

There exists no second order moment for the SαS probability distribution. Hence, the definitions of the measures of the noise level in a signal, Eqs. (40) and (41), are no longer satisfactory.

Different measures should be used in the case of SαS probability distributions. In [9] the fractional-order signal-to-noise ratio (*FSNR*) is proposed, which is defined as the ratio of the pth-order moments of the signal and the noise, where $0 < p < \alpha$, Eq. (42).

$$FSNR = 10 \log \frac{\mathrm{E}\{|s|^p\}}{\mathrm{E}\{|e|^p\}}, \quad 0 < p < \alpha \leqslant 2, \quad p \in \mathbb{R}. \tag{42}$$

The approximate comparison between *SNR* and *FSNR* is presented in Table 1.

A different noise-content measure for the asymptotical signals, like those in Fig. 2, called the instantaneous fractional-order signal-to-noise ratio $FSNR(t)$, Eq. (45), which is similar to the $SNR(t)$ proposed in [8], but is also suitable for noise with a SαS probability distribution, is proposed. The fractional-order moment $\mathrm{E}\{|s|^p\}$ for a linear single-degree-of-freedom system, Eq. (46), can be computed analytically. The zero phase can be taken into account, $\varphi = 0$, in Eq. (46) without any loss of generality and the fractional-order moment can be written as

$$\mathrm{E}\{|s|^p\} = \mathrm{E}\{|X \sin \omega_{od} t|^p\} = X^p C(p), \quad 0 \leqslant p, \; p \in \mathbb{R}, \tag{43}$$

where X denotes the slowly varying amplitude of the asymptotical signal, ω_{od} stands for the frequency of the response and $C(p)$ is the function dependent on the exponent p, Eq. (44), and is shown in Fig. 3.

$$C(p) = \frac{1}{\sqrt{\pi}} \frac{\Gamma((1+p)/2)}{\Gamma(1+p/2)}, \tag{44}$$

where $\Gamma(\cdot)$ stands for the gamma function. Thus, the $FSNR(t)$ for a linear single-degree-of-freedom system can be written as

$$FSNR(t) = 10 \log \left(C(p) \frac{X_0^p \mathrm{e}^{-p\delta\omega_o t}}{\mathrm{E}\{|e|^p\}} \right), \quad 0 < p < \alpha \leqslant 2, \; p \in \mathbb{R}, \tag{45}$$

where X_0 denotes the initial amplitude, δ stands for the viscous damping ratio and ω_o is the natural frequency.

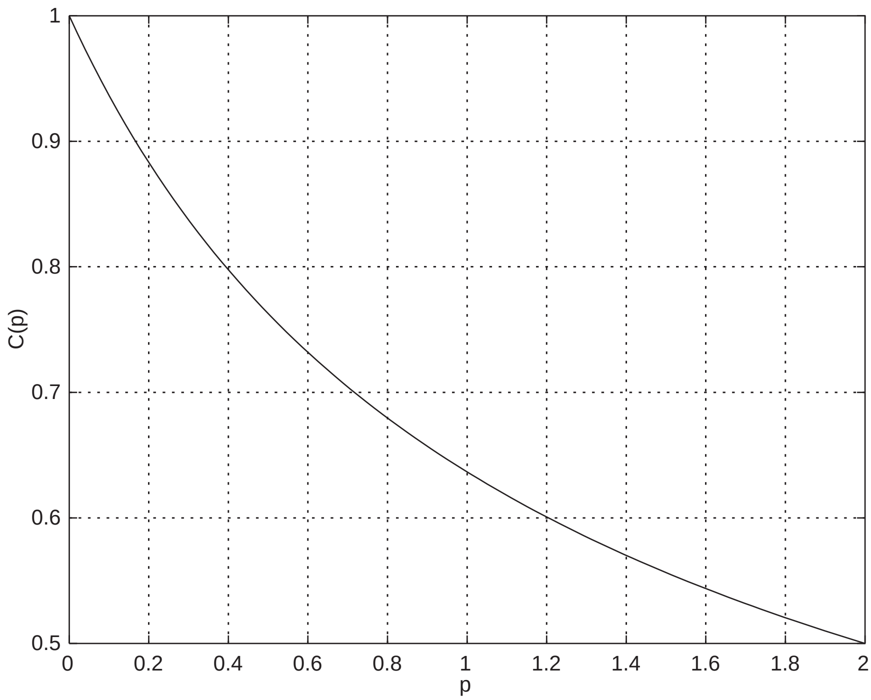

Fig. 3. The function $C(p)$.

The measures *FSNR* and *FSNR*(t) are suitable for signals with SαS-distributed noise as well as those with Gaussian and uniform noise. Both measures, *FSNR* and *FSNR*(t), broaden the applicability of the *SNR* and the *SNR*(t).

3. Numerical experiment

The numerical experiment was carried out using the linear single-degree-of-freedom system's response, Eq. (46):

$$x(t) = \mathrm{e}^{-\delta\omega_o t}[A\cos(\omega_{od}t) + B\sin(\omega_{od}t)] = \mathrm{e}^{-\delta\omega_o t}[X_0\sin(\omega_{od}t - \varphi)], \tag{46}$$

where δ stands for the viscous damping ratio, ω_o for the natural frequency, ω_{od} for the damped natural frequency, A and B are constants dependent on the initial conditions, X_0 stands for the initial amplitude and φ for the phase. Eq. (46) is rewritten as

$$x(t) = \mathrm{e}^{-Ct}[A\cos(Dt) + B\sin(Dt)], \tag{47}$$

which is more convenient for the parameter-identification procedure. The true values of the parameters are: $A = 1$, $B = -1$, $C = 0.1$ and $D = 0.99499$.

The *FSNR* is used as the signal's noise-content measure, because the comparison between the different methods of parameter identification is easier.

3.1. *The appropriateness of the multidimensional function minimization algorithms*

It was observed that the objective functions deduced with the help of the maximum-likelihood method are structured differently. Due to the heterogeneous structures of the objective functions, two algorithms of the gradient methods [4] (a numerically estimated gradient and an analytically computed one) and the downhill-simplex algorithm [4] (simplex [10]) were tested.

The focus was on the robustness on the initial guess of the model's parameters. It was found that the gradient methods converged only if the initial guess was near to the true values of the parameters, meaning that the methods are only successful when starting in the proximity of the minimum. In contrast, the downhill-simplex method searches for the minimum with a perseverance and speed that could not be matched by the gradient methods in these cases. This is consistent with the guidelines in Ref. [4].

Figs. 4 and 5 show the success of the downhill-simplex algorithm on a wide range of the initial values, A_o and B_o, of the parameters A and B. These values are proportional to the initial displacement and velocity, respectively. The convergence of the downhill-simplex method, when the initial values of the parameters C and D are taken as $C_o = 0$ and $D_o = 0$, which implies no additional effort to improve the initial values of parameters C and D, are presented in Fig. 4. The convergence of the downhill-simplex method is shown in Fig. 5, with the initial values of the parameters of $C_o = 0.01$ and $D_o = 1$ implying better initial values of the parameters. The value of D_o can be estimated from the power spectrum of the response and the value of C_o can be estimated, or at least the sign of the parameter C can be estimated, from the response, Fig. 2a. A small positive number for C_o, $C_o = 0.01$ was taken into account.

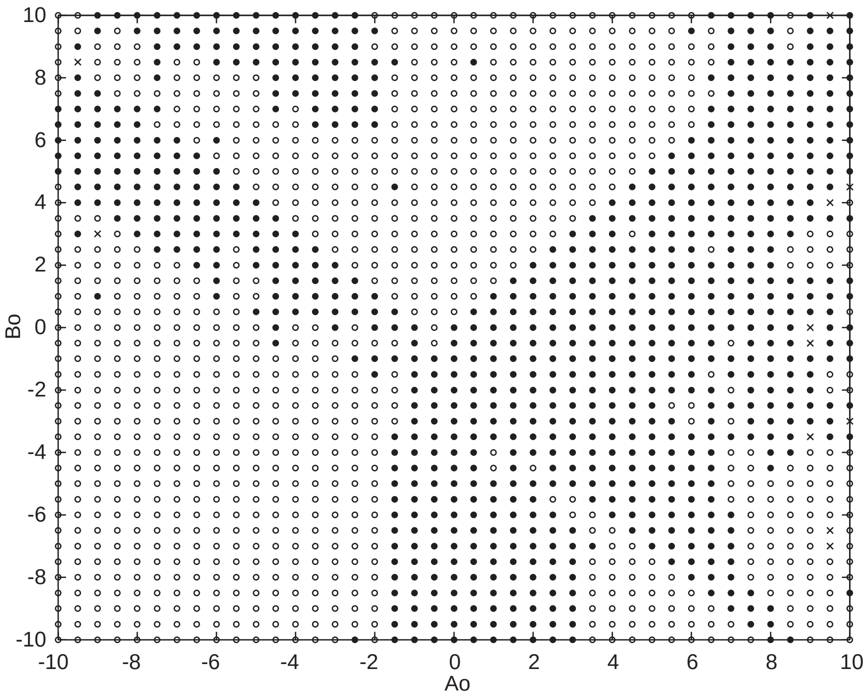

Fig. 4. The convergence of the downhill-simplex algorithm for different values of A_o and B_o and for $C_o = 0$ and $D_o = 0$. Legend: • convergence towards the proper minimum, ◦ convergence towards the other minima and × divergence.

3.2. Noise in the system's responses

The three different types of noise—uniform, Gaussian and Cauchy—were added with different levels of noise contamination measured with the *FSNR*, Eq. (42), to the system's responses. Each noise is made up of 1000 discrete data points as well as the responses.

The normalized noises and their histograms are presented in Fig. 6. The noises are normalized in a way that the maximum absolute value is set to unity. Each histogram consists of 50 bins. It is evident that the histograms only approximately correspond to their theoretical probability distributions. This is mainly due to the fact that the noise is of very limited length and the theoretical probability density functions are defined on an infinite number of data. Nevertheless, the differences in the different noises are visible in the time domain and in the histograms, Fig. 6.

Fig. 2 shows the system's response, to which the Gaussian noise of different levels is added. The response is dominant at $FSNR = 10$ dB, but at $FSNR = 0$ dB and $FSNR = -5$ dB the response is hidden by the noise.

The instantaneous fractional-order signal-to-noise ratio $FSNR(t)$ as shown in Fig. 7, where $FSNR = 0$ dB, can be used as a measure of the useful portion of the signal on which it is reasonable to perform the parameter identification, just like the $SNR(t)$ in Ref. [8]. The first 12 s would have been the best choice in our case, but the use of the *FSNR* measure on the whole response for the sake of an easy comparison of the different methods was preferred.

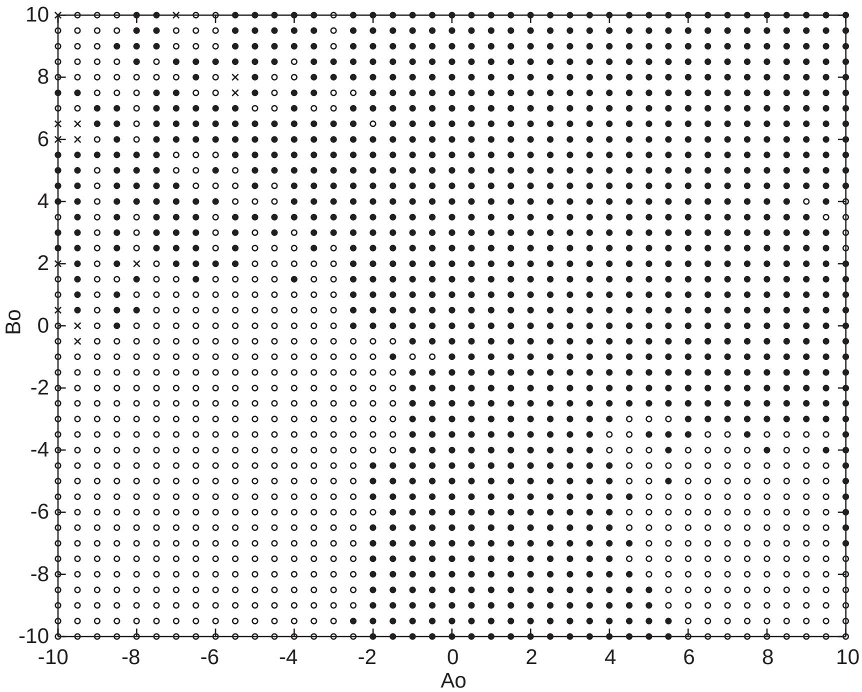

Fig. 5. The convergence of the downhill-simplex algorithm for different values of A_o and B_o and for $C_o = 0.01$ and $D_o = 1$. Legend: • convergence towards the proper minimum, ◦ convergence towards the other minima and × divergence.

3.3. Parameter identification on the responses with added uniform noise

The empirical analysis of the best choice of the generalized Cauchy p.d.f. exponent n is given in Table 2. A value for n between 3 and 5 is estimated as the best choice.

The parameter-identification results of the different methods are presented in Table 3 with abbreviations:

- LMSE—least-mean-squared-error method, Eq. (14);
- LAD—least-absolute-deviation method, Eq. (19);
- LPN—least-mean P-norm method (minimum dispersion method), Eq. (30);
- MLC—maximum-likelihood estimator for the Cauchy probability distribution, Eq. (25);
- MLU—maximum-likelihood estimator for the uniform probability distribution, Eq. (39);

It can be seen from Table 3 that the MLU method, which is theoretically best suited for the optimum parameter identification in presence of the uniform noise, is the most successful one, particularly at higher noise levels. It can also be seen that the LMSE and MLC methods produce the same results, which are better than the results obtained with the LAD and LPN methods.

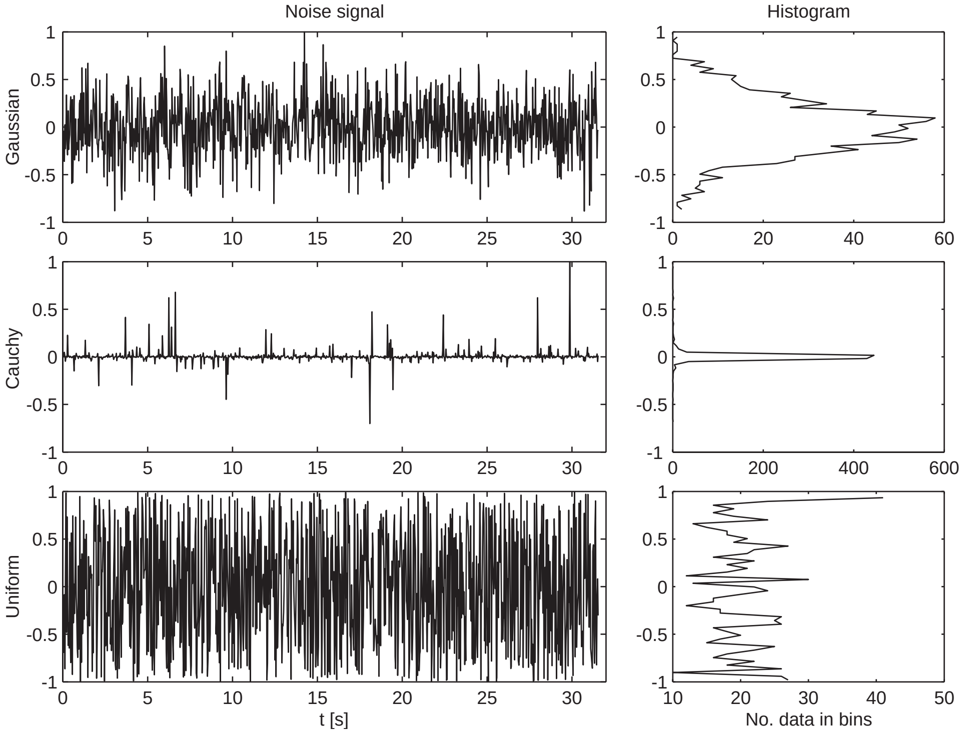

Fig. 6. Signals and histograms of Gaussian, Cauchy and uniform normalized noise.

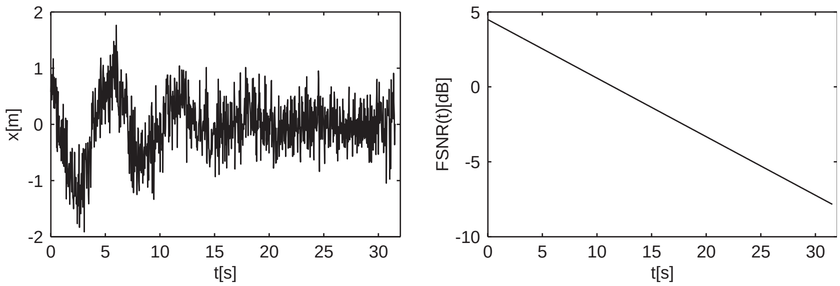

Fig. 7. The system's response with added Gaussian noise of $FSNR = 0$ dB and $FSNR(t)$ as function of time.

Table 2
Estimated values of the parameters with MLU method at different exponents n of the generalized Cauchy function and at uniform noise added to the response

$FSNR$ (dB)	Parameter	Value	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
0	A	1.0000	0.9681	0.9881	0.9988	1.0167	1.0371
	B	-1.0000	-1.0213	-1.0218	-1.0218	-1.0056	-0.7102
	C	0.1000	0.0958	0.0974	0.0985	0.0983	0.1315
	D	0.9950	0.9915	0.9926	0.9933	0.9942	1.0510

Table 3
Estimated parameter values at different uniform noise levels obtained with different methods

FSNR (dB)	Parameter	Value	Method				
			LMSE	LAD	$LPN_{p=0.9}$	MLC	$MLU_{n=4}$
10	*A*	1.0000	0.9917	0.9817	0.9808	0.9917	0.9998
	B	−1.0000	−1.0018	−0.9976	−0.9965	−1.0018	−1.0022
	C	0.1000	0.0992	0.0983	0.0983	0.0992	0.0998
	D	0.9950	0.9943	0.9935	0.9934	0.9943	0.9948
0	*A*	1.0000	0.9171	0.8224	0.8172	0.9171	0.9988
	B	−1.0000	−1.0173	−1.0296	−0.9762	−1.0173	−1.0215
	C	0.1000	0.0920	0.0878	0.0855	0.0920	0.0985
	D	0.9950	0.9884	0.9775	0.9778	0.9884	0.9933
−10	*A*	1.0000	0.2943	−0.2853	−0.2686	0.2943	0.9695
	B	−1.0000	−1.1347	−1.1622	−1.1119	−1.1347	−1.1855
	C	0.1000	0.0424	0.0272	0.0202	0.0424	0.0836
	D	0.9950	0.9422	0.8961	0.8905	0.9422	0.9808

Table 4
Estimated parameter values at different Gaussian noise levels obtained with different methods

FSNR (dB)	Parameter	Value	Method				
			LMSE	LAD	$LPN_{p=0.9}$	MLC	$MLU_{n=4}$
10	*A*	1.0000	1.0025	1.0007	1.0012	1.0025	1.0271
	B	−1.0000	−1.0020	−1.0092	−1.0124	−1.0020	−1.0162
	C	0.1000	0.1010	0.1016	0.1017	0.1010	0.1042
	D	0.9950	0.9951	0.9943	0.9939	0.9951	0.9947
0	*A*	1.0000	1.0240	1.0131	1.0161	1.0240	1.2662
	B	−1.0000	−1.0175	−1.1299	−1.1523	−1.0175	−1.1705
	C	0.1000	0.1103	0.1180	0.1179	0.1103	0.1406
	D	0.9950	0.9967	0.9841	0.9795	0.9967	0.9875
−5	*A*	1.0000	1.0603	0.9341	0.9341	1.0603	1.5799
	B	−1.0000	−1.0270	−1.4898	−1.5249	−1.0270	−1.5119
	C	0.1000	0.1298	0.1511	0.1493	0.1298	0.1828
	D	0.9950	1.0027	0.9290	0.9276	1.0027	0.9324

3.4. Parameter identification on the responses with added Gaussian noise

The results of the parameter identification on the responses with added Gaussian noise are presented in Table 4. The MLU method did badly in this case. The best results were obtained with the LMSE method, which is also the optimal one, in theory. The same results were obtained with

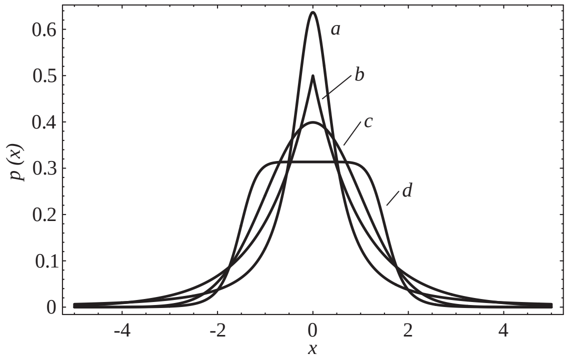

Fig. 8. Different probability density function.

the MLC method. The results of the LMSE, LAD and LPN methods are comparable for lower noise levels ($FSNR \geqslant 0$ dB), because the finite length of noise means that the histogram of the Gaussian noise, Fig. 6, is only nearly Gaussian. The other reason is that the shapes of the p.d.f.s of the methods, Fig. 8, are similar to the histogram's shape.

The marks in Fig. 8 stand for

$a\to$ Cauchy p.d.f., Eq. (21) considering $m = 0$ and $b = 1$,
$b\to$ Laplace p.d.f., Eq. (16) considering $\mu = 0$ and $\eta = 1$,
$c\to$ Gaussian p.d.f., Eq. (11) considering $\mu = 0$ and $\sigma = 1$ and
$d\to$ Generalized Cauchy p.d.f., Eq. (34) considering $m = 0$, $b = 1$ and $n = 4$.

3.5. Parameter identification on the responses with added Cauchy noise

The results of the parameter identification on the responses with added Cauchy noise are presented in Table 5. As expected, the best results were obtained with the MLC method. The results of the LMSE and MLC methods were not the same as they were for the responses with added uniform or Gaussian noise. The LPN method, originating from the minimum dispersion method of the symmetrical-α-stable distribution [6], is theoretically also suitable for the task. The results in Table 5 indicate that the MLC method is only slightly better than the LPN method. The results of the latter are comparable to the results of the LAD method, even though it is theoretically unsuitable for optimum parameter identification. The reason for the success of the LAD method lies in the finite length of the noise (only 1000 data points), which prevents us from reaching the theoretical Cauchy probability distribution, Fig. 6. The value α of the Cauchy noise, which should have been 1, was estimated using algorithm (29). The estimated value of α is slightly greater than 1, which is the consequence of the finite data length and according to Eq. (30) the LAD should be successful.

4. Conclusions

The mathematical theory of the function approximation, the maximum-likelihood method, the measures of optimality and the parameter identification were brought together and presented in this article.

Table 5
Estimated parameter values at different Cauchy noise levels obtained with different methods

FSNR (dB)	Parameter	Value	Method				
			LMSE	LAD	$LPN_{p=0.9}$	MLC	$MLU_{n=4}$
10	*A*	1.0000	1.0260	1.0054	1.0054	1.0048	1.1111
	B	−1.0000	−1.0127	−0.9998	−1.0000	−0.9999	−1.0289
	C	0.1000	0.0995	0.1003	0.1003	0.1003	0.0915
	D	0.9950	0.9955	0.9953	0.9953	0.9953	0.9966
0	*A*	1.0000	1.2592	1.0531	1.0537	1.0473	0.9460
	B	−1.0000	−1.0987	−0.9990	−1.0008	−0.9993	−0.9374
	C	0.1000	0.0946	0.1031	0.1035	0.1031	0.0812
	D	0.9950	1.0013	0.9984	0.9974	0.9980	1.0272
−10	*A*	1.0000	2.8092	1.5232	1.5450	1.5018	2.0437
	B	−1.0000	−0.7752	−1.0857	−1.1165	−1.0511	−1.6632
	C	0.1000	0.0365	0.1369	0.1450	0.1385	0.2690
	D	0.9950	1.0508	1.0152	1.0088	1.0173	1.1038

The probability density function (p.d.f.) of the uniform probability distribution is lacking its inverse function's values throughout in the interval $(-\infty, +\infty)$. The maximum-likelihood estimator for the p.d.f. of the uniform probability distribution is introduced on the basis of the generalized Cauchy p.d.f. It was shown that the generalized Cauchy function becomes the p.d.f. of the uniform probability distribution if subjected to the limitation process of $\lim_{n\to+\infty}$. Its advantages were empirically shown against all the other optimality measures presented in this paper.

It was shown that it is possible to generalize the measure of the instant noise level $SNR(t)$ to the $FSNR(t)$, which can be additionally used with symmetrical-α-stabile distributed noise.

The roots of the well-known measure of optimality, the least-mean-squared errors, are presented in this paper, as well as its optimality restricted only to Gaussian noise.

Parameter identification while different types of noise (uniform, Gaussian and Cauchy) were added to the response was used as an example of an optimization problem. It was shown that each type of the noise has its own maximum-likelihood estimator.

It was found out that the maximum-likelihood estimator based on the Cauchy p.d.f. (MLC) gives the same results as least-mean-squared-errors (LMSE) method, when taking into account the responses with added Gaussian and uniform noise. When dealing with Cauchy noise, the MLC method was better than the LMSE method.

The differences between the results of the maximum-likelihood estimators are small if the noise level is low ($FSNR > 10$). The optimality of the different estimators becomes important if the noise level is higher.

References

[1] N. Jakšić, M. Boltežar, An approach to parameter identification for a single-degree-of-freedom dynamical system based on short free acceleration response, *Journal of Sound and Vibration* 250 (3) (2002) 465–483.

[2] N. Jakšić, M. Boltežar, Parameter identification for single-degree-of-freedom dynamic systems = Prispevek k parameterski identifikaciji dinamičnih sistemov z eno prostostno stopnjo, *Journal of Mechanical Engineering = Strojniški vestnik* 48 (6) (2002) 302–317.
[3] G. Engeln-Müllges, F. Uhlig, *Numerical Algorithms with C*, Springer, Berlin, 1996.
[4] W. Press, B. Flannery, S. Teukolsky, W. Vetterling, *Numerical Recipes in Pascal, the Art of Scientific Computing*, Cambridge University Press, Cambridge, 1992.
[5] E. Weisstein, *Eric Weisstein's World of Mathematics*, CRC Press/Wolfram Research Boca Raton, FL/Champaign, IL. http://mathword.wolfram.com.
[6] M. Shao, C.L. Nikias, Signal processing with fractional lower order moments: stable processes and their applications, *Proceedings of the IEEE* 81 (7) (1993) 986–1009.
[7] P.G. Georgiou, P. Tsakalides, C. Kyriakakis, Alpha-stable modeling of noise and robust time-delay estimation in the presence of impulsive noise, *IEEE Transactions on Multimedia* 1 (3) (1999) 291–301.
[8] J. Slavič, I. Simonovski, M. Boltežar, Damping identification using a continuous wavelet transform: application to real data, *Journal of Sound and Vibration* 262 (2) (2003) 291–307.
[9] E.E. Kuruoğlu, P.W.J. Rayner, W.J. Fitzgerald, Least l_p-norm inpulsive noise cancellation with polynomal filters, *Signal Processing* 69 (1998) 1–14.
[10] J.A. Nelder, R. Mead, A simplex method for function minimization, *Computer Journal* 7 (1965) 308–313.